[0013] **FIG. 4A** is an exemplary category map that may be used for classifying a user's past search experience.

[0014] **FIG. 4B** is an exemplary data structure that may be used for storing category-based profiles for a plurality of users.

[0015] **FIG. 5** is an exemplary data structure that may be used for storing link-based profiles for a plurality of users.

[0016] **FIG. 6** is a flowchart illustrating paragraph sampling.

[0017] **FIG. 7A** is a flowchart illustrating context analysis.

[0018] **FIG. 7B** depicts a process of identifying important terms using context analysis.

[0019] **FIG. 8** illustrates a plurality of exemplary data structures that may be used for storing information about documents after term-based, category-based and/or link-based analyses, respectively.

[0020] **FIG. 9A** is a flowchart illustrating a personalized web search process according to one embodiment.

[0021] **FIG. 9B** is a flowchart illustrating a personalized web search process according to another embodiment.

[0022] The figures depict various embodiments of the present invention for purposes of illustration only. One skilled in the art will readily recognize from the following discussion that alternative embodiments of the illustrated and described structures, methods, and functions may be employed without departing from the principles of the invention.

DETAILED DESCRIPTION

[0023] System Overview

[0024] **FIG. 1** illustrates a system **100** in accordance with one embodiment of the present invention. System **100** comprises a front-end server **102**, a search engine **104** and associated content server **106**, a personalization server **108** and associated user profile server **110**, a content analysis module **112**, an advertisement server **114** and associated advertisement database **116**. During operation, a user accesses the system **100** via a conventional client **118** over a network (such as the Internet, not shown) operating on any type of client computing device, for example, executing a browser application. While only a single client **118** is shown, the system **100** supports large number of concurrent sessions with many clients. The system **100** operates on high performance server class computers; similarly the client device **118** can be any type of computing device. The details of the hardware aspects of server and client computers is well known to those of skill in the art and thus is not further described here.

[0025] The front-end server **102** is responsible for receiving a search query submitted by the client **119** along with some form of user ID that identifies either the user herself or the client device **118**. The front-end server **102** provides the query to the search engine **104**, which evaluates the query to retrieve a set of search results in accordance with the search query and returning the results to the front-end server **102**. The search engine **104** communicates with one or more content servers **106** and one or more user profile servers **108**. A content server **106** stores a large number of indexed documents indexed (and/or retrieved) from different websites. Alternately, or in addition, the content server **106** stores an index of documents stored on various websites. "Documents" are understood here to be any form of indexable content, including textual documents in any text or graphics format, images, video, audio, multimedia, presentations, and so forth. In one embodiment, each indexed document is assigned a rank or score using a link-based scoring function that takes into account an attribute associated with one or more links to the document. One example of a link-based scoring function is the page rank of a document. The page rank serves as a query independent measure of the document's importance. An exemplary form of page rank is described in U.S. Pat. No. 6,285,999 which is incorporated by reference. The search engine **104** communicates with one or more of the content servers **106** to select a plurality of documents that are relevant to user's search query. The search engine **104** assigns a score to each document based on the document's page rank, the text associated with the document, and the search query.

[0026] The personalization server **108** receives the search results from the search engine **104**, and the user ID from the front-end server **102**, and personalizes the results based on a profile of the user. The personalization server **108** communicates with the user profile server **110**, which stores a plurality of user profiles in a user profile database **110**. Each user profile includes information that identifies a user as well as describes the user's interests which can be used to refine the search results in response to the search queries submitted by this user. A user profile can be derived from a variety of different sources, such as the user's previous search experience, personal information, web pages associated with the user, and so forth. One embodiment for constructing the user's profile and using it to personalize search results is further described in the next section.

[0027] More specifically, the user profile server **108** receives the user ID from the front-end server **102**, and returns the associated profile to the personalization server **108**. The personalization server **108** personalizes the search results by rescoring and/or reranking the documents included there according to the user profile. The personalization server **108** provides the personalized search results back to the front-end server **102**.

[0028] The personalization server **108** also provides the personalized search results to the content analysis module **112**. The content analysis module **112** analyzes the content of the documents included in the search results (or a subset thereof), and derives a search profile that is descriptive of the documents. For example, the search profile can comprise key terms in the documents, topics or categories that describe the documents, website information from which the documents were retrieved, and so forth. Because the search profile is derived from the personalized search results, it reflects the personalization of the results, and thus the descriptive information preserves this personalization aspect.

[0029] The content analysis module **112** provides the search profile to the advertisement server **114**. The advertisement server **114** uses the search profile to select from the advertisement database **116** one or more advertisements for displaying in conjunction with the personalized search results. The selected personalized advertisements are provided to the front-end server **102**.